TRAINING & RESEARCH

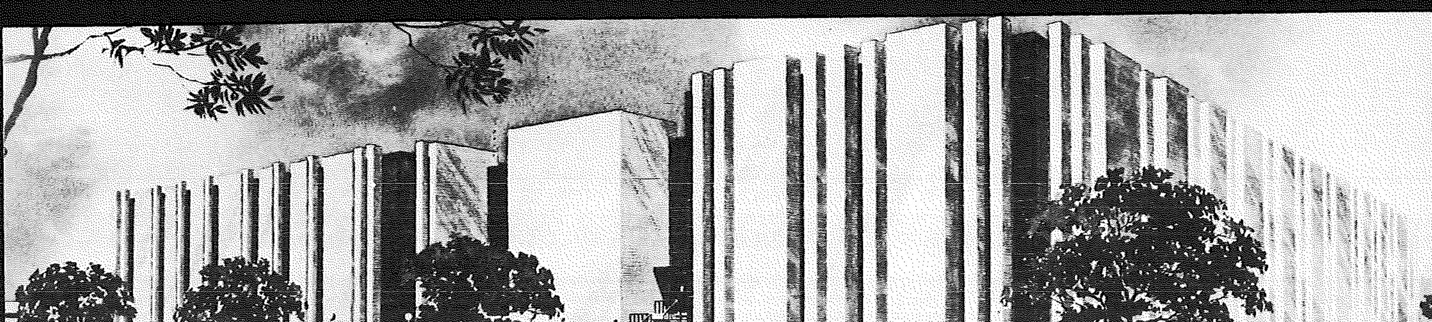
TRAINING & RESEARCH

TRAINING & RESEARCH

TRAINING & RESEARCH

TRAINING & RESEARCH

GRADUATE INSTITUTE OF STATISTICS



TEXAS A&M UNIVERSITY · COLLEGE STATION

N 70 28336

SEMI-ANNUAL STATUS REPORT

NASA CR 108409

THE EFFECTS OF SERIES TRUNCATION ON ESTIMATES AND THE TRANSFORMATION OF ERRORS

National Aeronautics and Space Administration

Grant #NGR 44-001-095

H. O. Hartley, Director
Institute of Statistics
Texas A&M University
College Station, Texas 77843

SEMI-ANNUAL STATUS REPORT

THE EFFECTS OF SERIES TRUNCATION ON ESTIMATES AND THE TRANSFORMATION OF ERRORS

I. Introduction.

The purpose of this research is to attack problems of predicting and measuring the statistical distributions of assembly performances in reliability testing and other engineering designs which employ performance characteristics of constituent parts. It was proposed to attack this problem by first investigating the effects of Taylor series truncation on the estimation of parameters in the expansion of the multivariate functions of interest; conditional expectations, Monte Carlo techniques, and standard estimation procedures were given as possibilities for this estimation. The second objective was to consider the statistical properties resulting from transformations on multivariate functions of random variables.

During the performance period outlined by this report, implementation of several procedures was begun. This research has taken form in parametric estimation using censored data from a particular statistical distribution, in implementation of an advanced Monte Carlo technique, and in the computer implementation of the maximum likelihood procedure for estimating the parameters in a general mixed analysis of variance model (originally proposed by Hartley and Rao (1967)). In addition to these attacks on the problem, estimation of variance components has begun using a technique similar to that of Goodman and Hartley (1958). A more detailed and specific explanation of each of the research procedures currently being employed is given in Section II.

The procedures outlined in the previous paragraph have all progressed sufficiently to allow technical reports to be issued shortly. There are however several other areas of research that are in an embryonic stage of development. Among these will be a thorough and intensive investigation into the statistical properties of parameter estimation given a function of many highly correlated random variables. It is desired to find either analytically or by computer simulation the properties of the estimates subject to several conditions:

(1) The specification of the joint multivariate distribution of the random variables of interest.

(2) An indication of the type of function of these random variables which can be considered to give rather "valid" estimates of the parameters of interest.

II. Current Research.

This section will include brief abstracts of the areas being currently investigated. The four abstracts outline research at such a stage of development that reports will be issued on each shortly. Each abstract will cover a specific area being investigated, the method of attack being used, and the membership of the research team being employed. The abstracts are given in the anticipated order in which the technical reports will be issued.

2.1. Lognormal Parameter Estimation from Censored Data.

This research topic deals with the aspects of maximum likelihood estimation of the three parameters of a lognormal distribution from censored (type 1) data. The existing techniques are discussed: Harter and Moore (1966), an iterative scheme for solving the simulteneous likelihood equations; Tiku (1968), a direct technique for solving modified likelihood equations. A further development due to Syler (1968) of Tiku's procedure is also considered, and a new type of estimation procedure based on a nonconvex programming algorithm is outlined. A quadratic scheme for determining the elements of the Fisher information matrix in each of the above situations, and hence the covariance matrix, is presented. The small sample biases of these estimates are determined, and Quenoille's method for eliminating the biases is extended to the multi-parameter case.

A Monte Carlo simulation is carried out in which the existing techniques are compared on the basis of expected biases and variance-covariance matrices for given parametric situations. In addition, a simulation is carried out in which random samples are generated from the three parameter lognormal distribution, and these parameters are estimated using the Hartley-Hocking convex programming algorithm. It should be noted that the likelihood function being considered here is not convex, but the procedure is carried out in an attempt to achieve a reasonable starting value for the other iterative procedure, to provide a "measure" of the non-convexity of the objective function, and to provide a contrast to a non-convex programming algorithm solution (Hartley, George, and LaMotte, (1969)) which when implemented on the computer should give a reasonable estimation technique. Briefly this non-convex programming procedure depends on a separability of the objective function into two additive functions, one of which is convex, the other which is non-convex. In this method the Hartley-Hocking convex programming algorithm is applied to optimizing the convex part of the objective function for a constant value of the non-convex portion. This is done for each iterate on the non-convex part, thus providing an "optimal" value of the convex procedure at each stage of the non-convex procedure. Thus if an iterative procedure can be developed so that the non-convex portion approaches some optimal point, then the optimality is guaranteed for the entire objective function. This procedure currently finds only a relative optimal point, that is, not necessarily a global optimal. Implementation of the procedure on the computer is not yet complete, however results on the existing techniques other than those of mathematical optimization will be reported and compared both analytically and computationally.

Team members: Dr. W. B. Smith, C. D. Zeis.

2.2. System Reliability Simulation by Stratified Monte Carlo: An Electronics Example.

This research topic deals with the problem of simulating on an electronic computer the expected results of complex machineries which have as inputs a large number of variable pieces of information. Thus, in order to properly illustrate the advantages of Stratified Monte Carlo over the Monte Carlo simulation procedure commonly used, the reliability of a complex piece of electronic equipment (in this case a portion of a television receiver) was simulated. The reasons for choosing this particular electronic device to study are:

(1) The device is sufficiently complicated to give a realistic illustration,

(2) this device was studied using a straightforward standard simulation technique (Bosinoff and Jacobs (1963)).

We show by using the modified technique of simulation that not only can considerable savings in computer time be realized, but also an increased precision of the resulting estimates can be given. It should be noted that the distribution as well as the parameters of this complex function of random variables can be estimated.

The Stratified Monte Carlo simulation technique (Suharto and Ringer (1967)) is being extended and implemented in the course of this research. Stratified Monte Carlo simulation is based fundamentally on the same principles as stratified sampling. That is, if we are observing a function of many random input variates, a standard procedure of simulation would be to generate the entire multivariate random data vector, calculate the value of the function of interest from this vector, record its value, compare it to design standards, and then repeat the procedure. Thus after many iterations of the procedure a empirical distribution function would be generated. The distribution function would estimate the underlying population distribution function of the random variable of interest. The greater the number of times this procedure is repeated the more accurate the empirical distribution function becomes with respect to the population distribution functions. Similar statements can be made about the population parameters as estimated.

Briefly, Stratified Monte Carlo simulation is a procedure which partitions the ranges of the input random variables into equal probability intervals. Then instead of sampling from the entire range of the input random variable as in conventional Monte Carlo, each combination of sub-ranges for each variate is sampled. The particular example considered has sixteen input variables, thus if each input variables range was divided into two equal probability sub-ranges, there would be 2^{16} possible equally weighted pseudo-random samples. Specific formulation of the reduction in variates and/or computer time are given in the forthcoming technical report. This report will include the extensions that have been made on the Stratified Monte Carlo procedure as well as this practical illustration of the procedure.

Team members: Dr. W. B. Smith, J. L. Oglesby.

2.3. Assessment of Sources of Variability by Maximum Likelihood Estimation: A Computer Program.

Hartley and Rao (1967) developed maximum likelihood estimates of the parameters in a general mixed analysis of variance model; the current research implements the procedure and extends it to the assessment of sources of variability. A report on this research will be submitted as a technical report shortly and will be published as a chapter in a book edited by T. A. Bancroft to be published soon by the Iowa State University Press in honor of George W. Snedecor.

A general mixed analysis of variance model involving both fixed and random factors with interaction can be given by

$$Y = X\alpha + U_1 b_1 + \ldots + U_c b_c + e, \qquad (1)$$

where X and U are matrices of known fixed numbers, α is a vector of unknown constants, b_1 is a vector of independent variables from $N(0, \sigma_i^2)$, and e is a vector of independent variables from $N(0, \sigma^2)$. Hartley and Rao give solutions to the likelihood equations resulting from repeated sampling on this model. These are the familiar least squares results

$$\begin{aligned} \tilde{\alpha} &= (X'H^{-1}X)^{-1}\,(X'H^{-1}y) \\ n\tilde{\sigma}^2 &= y'H^{-1}y - (X'H^{-1}y)(X'H^{-1}X)^{-1}(X'H^{-1}y), \end{aligned} \qquad (2)$$

where H is the variance-covariance matrix of y and

$$\gamma_i = \sigma_i^2/\sigma^2.$$

Using the method of steepest descent one would then be able to solve the likelihood equations for γ_i as well. A full explanation of the formulation is given in the paper by Hartley and Rao, however implementation of the procedure has just now been accomplished and is the subject of the report which will appear. Moreover, the procedure will be used and extended to the calculation of the variance-covariance matrix of the component of variance estimates. This extension will be carried out using statistical differentials (truncated Taylor series expansions).

Team members: Dr. H. O. Hartley, William Vaughn.

2.4. Confidence Regions for a Ratio of Variance in Variance Component Models.

Consider the model as described in equation (1). By letting

$$Z = \sum_{i=1}^{c} U_i b_i + e \tag{4}$$

and β be a vector describing the null hypothesis, then

$$y = X\alpha + (U_1 \vdots \ldots \vdots U_i)\beta + Z. \tag{5}$$

Note that y is distributed as a multivariate normal distribution with variance-covariance matrix $\sigma^2(H)$. A confidence region on the vector $\gamma' = (\gamma_1, \ldots \gamma_c)$ was given by Hartley and Rao.

Their procedure is now being extended to include one-way classification models with random effects, two-way classification models (random and mixed effects) and balanced incomplete designs with treatments fixed but blocks random. In the balanced cases the results are identical to the analysis of variance, however in the unbalanced cases the procedure does not always give the same results as in the analysis of variance. The procedure is being compared with that of Wald (1947).

Team members: Dr. H. O. Hartley, and A. Al-Barhawe.

III. Continuing and Future Research.

The topics reported on in Sections 2.1. and 2.2. are largely complete with the exception of applying mathematical optimization techniques to the functional described in 2.1. When results of this application are available an additional technical report will be submitted on the subject. The contents of the technical report to be submitted as outlined by Section 2.2. are largely completed.

Both Sections 2.3. and 2.4. represent "open ended" research topics. That is the results that are outlined in Sections 2.3. and 2.4. are in themselves considered sufficiently important to be technical reports. However each topic is of such scope that future research will be continued on these lines and additional technical reports are anticipated. Both topics constitute dissertation topics of the junior team members.

Additional research topics that will be considered under this contract are:

(1) A study of the properties of statistical differentials (truncated Taylor series),

(2) investigation into Bayesian estimation procedures in an attempt to achieve improved estimation,

(3) the application of the above results to quality control problems with many components,

(4) the extension of distribution free tolerance limits as investigated by Eisenhart, Hastay, and Willis to the problems of transformation of errors.

The above list of future research topics is of course not intended to exhaust all possibilities of potential research under this grant, but is intended only to indicate the general directions in which the majority of the research is headed. Additional research topics will stem from NASA sources and from extension of current research.

It is anticipated that whenever a technical report is written for submission under this grant that the research contained therein is of sufficient quality to be submitted to an appropriate scientific journal. Each of the first four topics (Sections 2.1. to 2.4.) will be so submitted.

V. References.

1. Bosinoff, I. and Jacobs, R. M. (1963). Using Transfer Functions in Reliability Prediction. Ninth National Symposium on Reliability and Quality Control, San Francisco.

2. Eisenhart, C., Hastay, M. V., and Willis, W. A. (1947). Techniques of Statistical Analysis, McGraw-Hill, New York.

3. Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators. J. Amer. Stat. Assoc., 53, 491-508.

4. Harter, H. L. and Moore, A. H. (1966). Local maximum likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. J. Amer. Stat. Assoc., 61, 842-851.

5. Hartley, H. O., George, M. and LaMotte, L. R. (1969). Mixed Convex and Non-Convex Programming, Project Themis Technical Report 21, Institute of Statistics, Texas A&M University.

6. Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 54, 93-108.

7. Ringer, L. J. and Suharto, S. (1967). Stratified Monte Carlo Sampling for Symmetric Statistics, Technical Report 9, ARO Contract DA-ARO-31-124-G793, Institute of Statistics, Texas A&M University.

8. Syler, G. W. (1968). Estimation of parameters in the lognormal distribution with censored samples. (unpublished thesis), Texas A&M University.

9. Tiku, M. L. (1968). Estimating the parameters of lognormal distribution from censored samples. J. Amer. Stat. Assoc., 63, 134-140.